Terms used **comparing rules** Found **18** of **132,857**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 18 of 18

Relevance scale

1 Passage-based Web text mining (poster session)
Thanaruk Theeramunkong
November 2000 **Proceedings of the fifth international workshop on on Information retrieval with Asian languages**
Full text available: pdf(173.11 KB) Additional Information: full citation, abstract, references

A large amount of textual information on the Web is very useful information resource. In the past, traditional text mining research treated a text document as a single piece of information. However, some Web documents are long and heterogeneous in their contents. This paper presents a new approach to apply the concept of a passage to Web text mining. A single Web text document is considered as several passages, instead of a single text. The effectiveness is investigated using real Thai Web do ...

Keywords: Thai Web documents, co-occurring, passage, text mining

2 Stack Machines and Classes of Nonnested Macro Languages
Joost Engelfriet, Erik Meineche Schmidt, Jan van Leeuwen
January 1980 **Journal of the ACM (JACM)**, Volume 27 Issue 1
Full text available: pdf(1.46 MB) Additional Information: full citation, references, citings, index terms

3 Common features of simulation based scheduling
F. Paul Wyman
December 1991 **Proceedings of the 23rd conference on Winter simulation**
Full text available: pdf(628.81 KB) Additional Information: full citation, references, citings, index terms

4 Grammar-like functional rules for representing query optimization alternatives
Guy M. Lohman
June 1988 **ACM SIGMOD Record , Proceedings of the 1988 ACM SIGMOD international c nference n Management of data**, Volume 17 Issue 3
Full text available: pdf(1.34 MB) Additional Information: full citation, abstract, references, citings, index terms

Extensible query optimization requires that the "repertoire" of alternative strategies for executing queries be represented as data, not embedded in the optimizer code. Recognizing that query optimizers are essentially expert systems, several researchers have suggested using strategy rules to transform query execution plans into alternative or better plans. Though extremely flexible, these systems can be very inefficient at any step in the

processing, many rules may be eligible ...

5 A general framework for formalizing UML with formal languages
William E. McUmber, Betty H. C. Cheng
July 2001 **Proceedings f the 23rd international conference on Software engineering**
Full text available: pdf(149.76 KB) Publisher Site Additional Information: full citation, abstract, references, citings, index terms

Informal and graphical modeling techniques enable developers to construct abstract representations of systems. Object-oriented modeling techniques further facilitate the development process. The Unified Modeling Language *(UML), an object-oriented modeling approach, could be broad enough in scope to represent a variety of domains and gain widespread use. Currently, UML comprises several different notations with no formal semantics attached to the individual diagrams. Therefore, it is ...*

***Keywords**: formal specifications, model checking, object-oriented modeling*

6 Verification of heuristic diagnostic knowledge by comparison with a causal/qualitative model
Graham F. Forsyth, Michael E. Larkin, Glen A. Wallace
June 1990 **Proceedings of the third international conference on Industrial and engineering applications of artificial intelligence and expert systems - Volume 2**
Full text available: pdf(576.40 KB) Additional Information: full citation, abstract, references, index terms

An approach to verify the knowledge base of a diagnostic expert system is described. An heuristic knowledge base collected from domain experts by interviews was analysed and the reasons for changes between versions were noted. The knowledge base was then compared with a small causal qualitative model of the device covered by the heuristic knowledge. Conclusions are drawn regarding the quality of the heuristic knowledge and indicate how it is planned to use the comparison of heuristic and ca ...

7 FreshML: programming with binders made simple
Mark R. Shinwell, Andrew M. Pitts, Murdoch J. Gabbay
August 2003 **ACM SIGPLAN Notices , Proceedings of the eighth ACM SIGPLAN international conference on Functional programming**, Volume 38 Issue 9
Full text available: pdf(187.31 KB) Additional Information: full citation, abstract, references, index terms

FreshML extends ML with elegant and practical constructs for declaring and manipulating syntactical data involving statically scoped binding operations. User-declared FreshML datatypes involving binders are concrete, in the sense that values of these types can be deconstructed by matching against patterns naming bound variables explicitly. This may have the computational effect of swapping bound names with freshly generated ones; previous work on FreshML used a complicated static type system inf ...

Keywords: alpha-conversion, metaprogramming, variable binding

8 Getting into a system: External-internal task mapping analysis
Thomas P. Moran
December 1983 **Proceedings of the SIGCHI conference on Human Fact rs in C mputing Systems**
Full text available: pdf(393.26 KB) Additional Information: full citation, abstract, references, citings, index terms

A task analysis technique, called ETIT analysis, is introduced. It is based on the idea that tasks in the external world must be reformulated into the internal concepts of a computer system before the system can be used. The analysis is in the form of a mapping between sets of external tasks and internal tasks. An example analysis of several text editing

systems is presented, and various properties of the systems are derived from the analysis. Further, it is shown how this analysis ...

9 Uniform self-stabilizing rings

J. E. Burns, J. Pachl

April 1989 **ACM Transactions on Programming Languages and Systems (TOPLAS)**, Volume 11 Issue 2

Full text available: pdf(1.12 MB) Additional Information: full citation, abstract, references, citings, index terms, review

A self-stabilizing system has the property that, no matter how it is perturbed, it eventually returns to a legitimate configuration. Dijkstra originally introduced the self-stabilization problem and gave several solutions for a ring of processors in his 1974 Communications of the ACM paper. His solutions use a distinguished processor in the ring, which effectively acts as a controlling element to drive the system toward stability. Dijkstra has observed that ...

10 A functional approach to integrating database and expert systems

Tore Risch, René Reboh, Peter E. Hart, Richard O. Duda

December 1988 **Communications of the ACM**, Volume 31 Issue 12

Full text available: pdf(1.67 MB) Additional Information: full citation, abstract, references, citings, index terms, review

A new system architecture shares certain characteristics with database systems, expert systems, functional programming languages, and spreadsheet systems, but is very different from any of these.

11 Representation results for defeasible logic

Grigoris Antoniou, David Billington, Guido Governatori, Michael J. Maher

April 2001 **ACM Transactions on Computational Logic (TOCL)**, Volume 2 Issue 2

Full text available: pdf(228.29 KB) Additional Information: full citation, abstract, references, citings, index terms, review

The importance of transformations and normal forms in logic programming, and generally in computer science, is well documented. This paper investigates transformations and normal forms in the context of Defeasible Logic, a simple but efficient formalism for nonmonotonic reasoning based on rules and priorities. The transformations described in this paper have two main benefits: on one hand they can be used as a theoretical tool that leads to a deeper understanding of the formalism, and on th ...

Keywords: defeasible logic, normal forms, transformations

12 DIAGRAM: a grammar for dialogues

Jane J. Robinson

January 1982 **Communications of the ACM**, Volume 25 Issue 1

Full text available: pdf(2.11 MB) Additional Information: full citation, abstract, references, citings, index terms

An explanatory overview is given of DIAGRAM, a large and complex grammar used in an artificial intelligence system for interpreting English dialogue. DIAGRAM is an augmented phrase-structure grammar with rule procedures that allow phrases to inherit attributes from their constituents and to acquire attributes from the larger phrases in which they themselves are constituents. These attributes are used to set context-sensitive constraints on the acceptance of an analysis. Constraints can be i ...

Keyw rds: annotations, attribute inheritance, augmented rules, contextual constraints, dialogue, likelihoods, metarules, phrase-structure grammar, transformations

13 Context-sensitive parsing

William A. Woods
July 1970 **Communications f the ACM**, Volume 13 Issue 7

Full text available: pdf(1.13 MB) Additional Information: full citation, abstract, references

This paper presents a canonical form for context-sensitive derivations and a parsing algorithm which finds each context-sensitive analysis once and only once. The amount of memory required by the algorithm is essentially no more than that required to store a single complete derivation. In addition, a modified version of the basic algorithm is presented which blocks infinite analyses for grammars which contain loops. The algorithm is also compared with several previous parsers for context-se ...

Keywords: context-sensitive grammars, context-sensitive parsing, formal grammars, formal language theory, parsing, parsing algorithms, recognition algorithms

14 A declarative approach to business rules in contracts: courteous logic programs in XML

Benjamin N. Grosof, Yannis Labrou, Hoi Y. Chan
November 1999 **Proceedings of the 1st ACM conference on Electronic commerce**

Full text available: pdf(140.64 KB) Additional Information: full citation, references, citings, index terms

15 A flexible interactive control structure for rule-based systems

S. Srinivasan, Pradip Dey, Yoichi Hayashi
February 1988 **Proceedings of the 1988 ACM sixteenth annual conference on Computer science**

Full text available: pdf(750.12 KB) Additional Information: full citation, abstract, references, citings, index terms

Flexibility in control mechanism will allow solutions of a much wider range of problems with the expert system technology than currently possible. In order to provide flexibility in control mechanism deviations from the standard fixed control (recognize-act cycle) should be allowed. As a first step toward achieving this we develop a flexible interactive backtracking strategy that can deviate significantly from the fixed control structure of rule-based systems. This paper describes a general ...

16 Mining the most interesting rules

Roberto J. Bayardo, Rakesh Agrawal
August 1999 **Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining**

Full text available: pdf(1.29 MB) Additional Information: full citation, references, citings, index terms

17 AnnoDomini: from type theory to Year 2000 conversion tool

Peter Harry Eidorff, Fritz Henglein, Christian Mossin, Henning Niss, Morten Heine Sørensen, Mads Tofte
January 1999 **Proceedings of the 26th ACM SIGPLAN-SIGACT symposium on Principles of programming languages**

Full text available: pdf(1.60 MB) Additional Information: full citation, references, citings, index terms

18 Analysis of rule sets generated by the CN2, ID3, and multiple convergence symbolic learning methods

Elizabeth M. Boll, Daniel C. St. Clair
February 1995 **Proceedings of the 1995 ACM 23rd annual c nference on C mputer science**

Full text available: pdf(898.31 KB) Additional Information: full citation, references, index terms

Results 1 - 18 of 18

Your search matched **4** of **1037503** documents.
A maximum of **500** results are displayed, **15** to a page, sorted by **Relevance** in **Descending** order.

Refine This Search:
You may refine your search by editing the current search expression or entering a new one in the text box.

'comparing rules' Search

☐ Check to search within this result set

Results Key:
JNL = Journal or Magazine **CNF** = Conference **STD** = Standard

---

1 **A fuzzy expert system for fault detection in statistical process control of industrial processes**
El-Shal, S.M.; Morris, A.S.;
Systems, Man and Cybernetics, Part C, IEEE Transactions on , Volume: 30 , Issue: 2 , May 2000
Pages:281 - 289

[Abstract] [PDF Full-Text (236 KB)] **IEEE JNL**

---

2 **Multi-script handwriting recognition with FOHDEL**
Malaviya, A.; Leja, C.; Peters, L.;
Fuzzy Information Processing Society, 1996. NAFIPS. 1996 Biennial Conference of the North American , 19-22 June 1996
Pages:147 - 151

[Abstract] [PDF Full-Text (412 KB)] **IEEE CNF**

---

3 **Grapheme-to-phone using finite-state transducers**
Caseiro, D.; Trancoso, L.; Oliveira, L.; Viana, C.;
Speech Synthesis, 2002. Proceedings of 2002 IEEE Workshop on , 11-13 Sept. 2002
Pages:215 - 218

[Abstract] [PDF Full-Text (510 KB)] **IEEE CNF**

---

4 **Augmenting UML with fact-orientation**
Halpin, T.;
System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on , 3-6 Jan. 2001
Pages:10 pp.

[Abstract] [PDF Full-Text (192 KB)] **IEEE CNF**

Term used **administration**

Found **9** of **60 searched out of 60.**

Sort results by: relevance

Display results: expanded form

Save results to a Binder

Search Tips

Open results in a new window

Try an Advanced Search

Try this search in The ACM Guide

Results 1 - 9 of 9

Relevance scale

1 A graph-based formalism for RBAC

Manuel Koch, Luigi V. Mancini, Francesco Parisi-Presicce

August 2002 **ACM Transactions on Information and System Security (TISSEC)**, Volume 5 Issue 3

Full text available: pdf(819.71 KB) Additional Information: full citation, abstract, references, citings, index terms, review

Role-Based Access Control (RBAC) is supported directly or in a closely related form, by a number of products. This article presents a formalization of RBAC using graph transformations that is a graphical specification technique based on a generalization of classical string grammars to nonlinear structures. The proposed formalization provides an intuitive description for the manipulation of graph structures as they occur in information systems access control and a precise specification of static ...

Keywords: Access control in information systems, correctness, decentralized administration, graph transformations, permission management, role-based access control

2 Automatic Subject Recognition in Scientific Papers: An Empirical Study

John O'Connor

October 1965 **Journal of the ACM (JACM)**, Volume 12 Issue 4

Full text available: pdf(1.65 MB) Additional Information: full citation, references, citings, index terms

3 Fast detection of communication patterns in distributed executions

Thomas Kunz, Michiel F. H. Seuren

November 1997 **Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research**

Full text available: pdf(4.21 MB) Additional Information: full citation, abstract, references, index terms

Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...

4 Computer applications in health care (CAHC): Compression of mammograms for medical practice

Artur Przelaskowski

March 2004 **Proceedings of the 2004 ACM symposium n Applied computing**

Full text available: pdf(244.15 KB) Additional Information: full citation, abstract, references, index terms

This paper considers effective compression methods for mammogram storing and interchange. A controversy problem of irreversible compression of medical images is studied in clinical tests to check usefulness and possibility of acceptance of wavelet-based compression for clinical applications. Diagnostic accuracy is measured in abnormality detection tests with ROC-based analysis, and by subjective rating of diagnostically important image features affecting lesion symptoms and image ordering accord ...

Keywords: diagnostic accuracy evaluation, image compression

5 Mobile objects in distributed Oz

Peter Van Roy, Seif Haridi, Per Brand, Gert Smolka, Michael Mehl, Ralf Scheidhauer

September 1997 **ACM Transactions on Programming Languages and Systems (TOPLAS)**, Volume 19 Issue 5

Full text available: pdf(484.83 KB) Additional Information: full citation, abstract, references, citings, index terms

Some of the most difficult questions to answer when designing a distributed application are related to mobility: what information to transfer between sites and when and how to transfer it. Network-transparent distribution, the property that a program's behavior is independent of how it is partitioned among sites, does not directly address these questions. Therefore we propose to extend all language entities with a network behavior that enables efficient distributed programm ...

Keywords: latency tolerance, mobile objects, network transparency

6 Formal semantics of APL: a review of initial findings

Phil Chastney

June 2002 **ACM SIGAPL APL Quote Quad , Proceedings of the 2002 conference on APL: array processing languages: lore, problems, and applications**, Volume 32 Issue 4

Full text available: pdf(83.89 KB) Additional Information: full citation, references

7 On the specification and evolution of access control policies

M. Koch, L. V. Mancini, F. Parisi-Presicce

May 2001 **Proceedings of the sixth ACM symposium on Access control models and technologies**

Full text available: pdf(240.60 KB) Additional Information: full citation, abstract, references, citings, index terms

A uniform and precise framework for the specification of access control policies is proposed. The uniform framework allows the detailed comparison of different policy models, the precise description of the evolution of a policy, and an accurate analysis of the interaction between policies and of the behavior of their integration. The evolution and integration of policies are illustrated using a Discretionary Access Control policy and a Lattice Based Access Control policy. The framework is b ...

Keywords: graph transformation systems, methodology, specification

8 A region coloring technique for scene analysis

James P. Strong, Azriel Rosenfeld

April 1973 **Communicati ns of the ACM**, Volume 16 Issue 4

Full text available: pdf(1.01 MB) Additional Information: full citation, abstract, references, citings

A method of converting a picture into a "cartoon" or "map" whose regions correspond to differently textured regions is described. Texture edges in the picture are detected, and solid regions surrounded by these (usually broken) edges are "colored in" using a propagation

process. The resulting map is cleaned by comparing the region colors with the textures of the corresponding regions in the picture, and also by merging some regions with others according t ...

Keyw rds: edge detection, picture processing, scene analysis

9 A new framework for elimination-based data flow analysis using DJ graphs

Vugranam C. Sreedhar, Guang R. Gao, Yong-Fong Lee
March 1998 **ACM Transactions on Programming Languages and Systems (TOPLAS)**, Volume 20 Issue 2
Full text available: pdf(631.44 KB) Additional Information: full citation, references, citings, index terms

Keywords: DJ graphs, Tarjan's interval, exhaustive and incremental data flow analysis, irreducible flowgraphs, reducible flowgraphs

Results 1 - 9 of 9

US-PAT-NO: 6606625

DOCUMENT-IDENTIFIER: US 6606625 B1

TITLE: Wrapper induction by hierarchical data analysis

DATE-ISSUED: August 12, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Muslea; Ion	Culver City	CA	N/A	N/A
Minton; Steven	El Segundo	CA	N/A	N/A
Knoblock; Craig A.	El Segundo	CA	N/A	N/A

US-CL-CURRENT: 707/6, 707/10

ABSTRACT:

An inductive algorithm, denominated STALKER, generating high accuracy extraction rules based on user-labeled training examples. With the tremendous amount of information that becomes available on the Web on a daily basis, the ability to quickly develop information agents has become a crucial problem. A vital component of any Web-based information agent is a set of wrappers that can extract the relevant data from semistructured information sources. The novel approach to wrapped induction provided herein is based on the idea of hierarchical information extraction, which turns the hard problem of extracting data from an arbitrarily complex document into a series of easier extraction tasks. Labeling the training data represents the major bottleneck in using wrapper induction techniques, and experimental results show that STALKER performs significantly better than other approaches; on one hand, STALKER requires up to two orders of magnitude fewer examples than other algorithms, while on the other hand it can handle information sources that could not be wrapped by prior techniques. STALKER uses an embedded catalog formalism to parse the information source and render a predictable structure from which information may be extracted or by which such information extraction may be facilitated and made easier.

13 Claims, 20 Drawing figures

Exemplary Claim Number: 1

Number of Drawing Sheets: 18

---------- KWIC ---------

Detailed Description Text - DETX (2):

The detailed description set forth below in connection with the appended drawings is intended as a description of presently-preferred embodiments of the invention and is not intended to represent the only forms in which the present invention may be constructed and/or utilized. The description sets forth the functions and the sequence of steps for constructing and operating the invention in connection with the illustrated embodiments. However, it is to be understood that the same or equivalent functions and sequences may be accomplished by different embodiments that are also intended to be **encompassed** within the spirit and scope of the invention.

Claims Text - CLTX (1):

1. A method for inducing or learning extraction rules for extracting data from a collection of data records, the steps comprising: providing examples of the collection of data records to provide an example set; indicating desired information in said example set; providing a rule list, said rule list initially empty; learning a rule based upon said example set and returning a new learned rule, said step of learning a new rule base upon said example including designating a seed example, creating rule candidates based upon said seed example to provide a candidate set, said rule candidates created by creating a two-state (2-state) landmark automaton for each seed token t that ends a prefix immediately preceding desired information, by creating a 2-state landmark automaton for each wildcard matching each of said seed tokens t, and by collecting all said 2-state landmark automations to provide said candidate set, refining said candidate set to provide a refined candidate set, and returning said refined candidate set; adding said new learned rule to said rule list removing all examples covered by said new learned rule from said example set to provide a revised example set; defining said example set as said revised example set; repeating the steps of learning a **new rule, adding said new learned rule, and removing** all covered examples until said example set is empty; and returning said rule list; whereby said rule list provides a set of rules by which desired information may be identified for extraction from said collection of data records and other data records similar to said examples.

Claims Text - CLTX (9):

9. A method for inducing or learning extraction rules for extracting data from a collection of data records, the steps comprising: providing examples of the collection of data records to provide an example set; indicating desired information in said example set; providing a rule list, said rule list initially empty; learning a rule based upon said example set and returning a new learned rule by designating a seed example, said seed example being a shortest example in said example set having a fewest number of tokens, creating rule candidates based upon said seed example to provide a candidate set, by refining said candidate set to provide a refined candidate set, and by returning said refined candidate set; said rule candidates created by creating a two-state (2-state) landmark automaton for each seed token t that ends a prefix immediately preceding desired information, by creating a 2-state landmark automaton for each wildcard matching each of said seed tokens t, and by collecting all said 2-state landmark automations to provide said candidate set; said candidate set refined by refining said candidate set to provide a new candidate set, by determining if a perfect solution has been achieved in said new candidate set, by, if necessary, repeating said refining and determining steps upon said new candidate as said candidate set until a perfect solution has been achieved, and by returning a resulting candidate set as said refined candidate set; said candidate set further refined by determining a best refined candidate rule from said candidate set, creating a token-based rule for each token t in said seed example that precedes a landmark l of said best refined candidate rule present in said seed example, said token-based rule adding a landmark automata based on said token to said best refined candidate rule and collecting each of said token-based rules in a token rule set, by creating a wildcard-based rule for each rule in said token rule set by substituting all valid wildcards for each token t in said token-based rule by adding a landmark automata based on each of said wildcards to said best refined candidate rule and collecting each of said wildcard-based rules in a wildcard rule set, by eliminating duplicates in said token rule set and said wildcard rule set, by repeating said steps of creating said token rule set and said wildcard rule set for each landmark in said best refined candidate rule, and by collecting all rules in a topology refinement rule set; said step of determining a best refined candidate including selecting best refined candidates based on candidates selected from the group consisting of: candidates providing larger coverage, candidates providing more early matches,

candidates providing more failed matches, candidates having fewer wildcards, candidates having shorter unconsumed prefixes, candidates having fewer tokens in SkipUntil( ) statements, candidates having longer end-landmarks; said candidate set further refined by determining a best refined candidate rule from said candidate set, by providing a sequence of consecutive tokens present in said seed example, by matching a first landmark in said best refined candidate rule with said sequence to provide a match, by creating a pre-landmark token rule by adding a token in said sequence immediately preceding said match, said pre-landmark token rule being a landmark automata based on the combination of said preceding token and said first landmark, by creating a post-landmark token rule by adding a token in said sequence immediately following said match, said post-landmark token rule being a landmark automata based on a combination of said following token and said first landmark, by creating pre-landmark wildcard-based rules by substituting all valid wildcards for said preceding token in said pre-landmark token rule by adding a landmark automata based on a combination of each of said wildcards and said first landmark and collecting each of said pre-landmark wildcard-based rules in a pre-landmark wildcard rule set, by creating post-landmark wildcard-based rules by substituting all valid wildcards for said following token in said post-landmark token rule by adding a landmark automata based on a combination of each of said wildcards and said first landmark and collecting each of said post-landmark wildcard-based rules in a post-landmark wildcard rule set, by repeating said steps of creating said pre-landmark token rule, creating said post-landmark token rule set, creating said pre-landmark wildcard-based rule set, and creating said post-landmark wildcard rule set for each landmark in said best refined candidate rule matching a sequence of consecutive tokens present in said seed example, and by collecting all rules so generated in a landmark refinement rule set; said step of determining if a perfect solution has been achieved including determining a current best solution from a union of a prior best solution with said candidate set and by selecting best solution candidates based on candidates selected from the group consisting of: candidates having more correct matches, candidates having more failures to match, candidates having fewer tokens in SkipUntil( ) statements, candidates having fewer wildcards, candidates having longer end-landmarks, and candidates having shorter unconsumed prefixes; adding said new learned rule to said rule list; removing all examples covered by said new learned rule from said example set to provide a revised example set; defining said example set as said revised example set; repeating the steps of learning a **new rule, adding said new learned rule, and removing** all covered examples until said example set is empty; and returning said rule list; whereby said rule list provides a set of rules by which desired information may be identified for extraction from said collection of data records and other data records similar to said examples.

Claims Text - CLTX (10):

10. A method for inducing or learning extraction rules for extracting data from a collection of data records, the steps comprising: providing examples of the collection of data records to provide an on each of said wildcards to said best refined candidate rule and collecting each of said wildcard-based rules in a wildcard rule set, eliminating duplicates in said token rule set and said wildcard rule set, repeating said steps of creating said token rule set and said wildcard rule set for each landmark in said best refined candidate rule, and collecting all rules in a topology refinement rule set; adding said new learned rule to said rule list; removing all examples covered by said new learned rule from said example set to provide a revised example set; defining said example set as said revised example set; repeating the steps of learning a **new rule, adding said new learned rule, and removing** all covered examples until said example set is empty; and returning said rule list; whereby said rule list provides a set of rules by which desired information may be identified for extraction from said collection of data records and other data records similar to said examples.

Claims Text - CLTX (12):

12. A method for inducing or learning extraction rules for extracting data from a collection of data records, the steps comprising: providing examples of the collection of data records to provide an example set; indicating desired information in said example set; providing a rule list, said rule list initially empty; learning a rule based upon said example set and returning a new learned rule; said step of learning a new rule including designating a seed example, creating rule candidates based upon said seed example to provide a candidate set, refining said candidate set to provide a refined candidate set, and returning said refined candidate set; said step of refining said candidate set including refining said candidate set to provide a new candidate set, determining if a perfect solution has been achieved in said candidate set, if necessary, repeating said refining and determining steps upon said new candidate set in place of said candidate set until a perfect solution has been achieved, if necessary, and returning said new candidate set; said step of refining said candidate set also including determining a best refined candidate rule from said candidate set, providing a sequence of removing all examples covered by said new learned rule from said example set to provide a revised example set; defining said example set as said revised example set; repeating the steps of learning a **new rule, adding said new learned rule, and removing** all covered examples until said example set is empty; and returning said rule list; whereby said rule list provides a set of rules by which desired information may be identified for extraction from said collection of data records and other data records similar to said examples.

Claims Text - CLTX (13):

13. A method for inducing or learning extraction rules for extracting data from a collection of data records, the steps comprising: providing examples of the collection of data records to provide an example set; indicating desired information in said example set; providing a rule list, said rule list initially empty; learning a rule based upon said example set and returning a new learned rule; said step of learning a new rule including designating a seed example, creating rule candidates based upon said seed example to provide a candidate set, refining said candidate set to provide a refined candidate set, and returning said refined candidate set; said step of refining said candidate set including refining said candidate set to provide a new candidate set, determining if a perfect solution has been achieved in said candidate set, if necessary, repeating said refining and determining steps upon said new candidate set in place of said candidate set until a perfect solution has been achieved, if necessary, and returning said new candidate set; said step of determining if a perfect solution has been achieved including determining a current best solution from a union of a prior best solution with said candidate set; said step of determining a current best solution including selecting best solution candidates based on candidates selected from the group consisting of: candidates having more correct matches, candidates having more failures to match, candidates having fewer tokens in SkipUntil( ) statements, candidates having fewer wildcards, candidates having longer end-landmarks, and candidates having shorter unconsumed prefixes; adding said new learned rule to said rule list removing all examples covered by said new learned rule from said example set to provide a revised example set; defining said example set as said revised example set; repeating the steps of learning a **new rule, adding said new learned rule, and removing** all covered examples until said example set is empty; and returning said rule list; whereby said rule list provides a set of rules by which desired information may be identified for extraction from said collection of data records and other data records similar to said examples.

	Type	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition
1	BRS	1723	"new rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:23		
2	BRS	493	"existing rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:24		
3	BRS	252	"new rules" and "existing rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:24		
4	BRS	198	"new rules" same "existing rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:24		
5	BRS	4	eliminat$6 near8 "existing rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:27		
6	BRS	23	eliminat$6 near8 "new rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:25		
7	BRS	7	(eliminat$6 near8 "new rules") and @py<2001	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:26		
8	BRS	13	remov$3 near8 "existing rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:28		
9	BRS	39	remov$3 near8 "new rules"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:28		
10	BRS	17	((remov$3 near8 "existing rules") or (remov$3 near8 "new rules")) and encompass$4	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/15 16:28		

	Errors
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0